Derivations of some Stability Conditions Using von Neumann Analysis

L. Davidson
Dept. Thermo and Fluid Dynamics
Chalmers University of Technology
SE-412 96 Gothenburg

February 16, 2001

1 Introduction

In this short note Neumann stability analysis is used to show:

- that a central differencing scheme is unstable;
- that both a second-order derivative and a fourth-order derivative are dissipative; this is true for all the nth-order derivative, where n is even;
- that the Crank-Nicolson scheme is stable.

In the final section, we also show that the Gauss-Seidel solver converges if $|a_P| \geq \sum |a_{nb}|$

2 Central differencing

We want to solve the unsteady convection equation (zero viscosity)

$$\frac{\partial T}{\partial t} + a\frac{\partial T}{\partial x} = 0. \tag{1}$$

Discretizing this equation on an equidistant mesh, using explicit discretization in time and central differencing in space gives

$$\frac{T_i^{n+1} - T_i^n}{\Delta t} = -a\frac{T_{i+1}^n - T_{i-1}^n}{2\Delta x}. \tag{2}$$

Note that explicit time integration is similar to solving a steady problem in 2D or 3D, because in 2D or 3D the iterative solver is always explicit to a certain degree. Denote the exact solution to Eq. 2 by $\bar{T}_i^n$ and the actual computed (i.e. approximate) solution by T_i^n. They are related as

$$T_i^n = \bar{T}_i^n + \varepsilon_i^n \tag{3}$$

where ε^n is the error at time level n, which may be due to, for example, round-off errors. Insert Eq. 3 into Eq. 2 so that

$$\frac{\bar{T}_i^{n+1} - \bar{T}_i^n}{\Delta t} + \frac{\varepsilon_i^{n+1} - \varepsilon_i^n}{\Delta t} = -a\frac{\bar{T}_{i+1}^n - \bar{T}_{i-1}^n}{2\Delta x} - a\frac{\varepsilon_{i+1}^n - \varepsilon_{i-1}^n}{2\Delta x}. \tag{4}$$

Since $\bar{T}_i$ exactly satisfies Eq. 2 we get an equation for the error, i.e.

$$\frac{\varepsilon_i^{n+1} - \varepsilon_i^n}{\Delta t} = -a\frac{\varepsilon_{i+1}^n - \varepsilon_{i-1}^n}{2\Delta x}. \tag{5}$$

The error ε can be represented as a Fourier series

$$\varepsilon_i^n = \sum_{m=-N}^{N} E_m^n e^{\jmath k_m x} = \sum_{m=-N}^{N} E_m^n e^{\jmath k_m i \Delta x} \tag{6}$$

where E_m^n is the amplitude and $\jmath = \sqrt{-1}$. The computational domain is $[0, L]$. The wave number is denoted by k which is defined as the period length 2π divided by the wave length λ (see Fig. 1), i.e.

$$k = \frac{2\pi}{\lambda}. \tag{7}$$

The minimum wave number corresponds to the largest wave length (the entire computation domain), and the largest wave number corresponds to the smallest wave length (the cell size), see Fig. 1, i.e.

$$\begin{aligned} k_{min} &= \frac{2\pi}{\lambda_{max}} = \frac{\pi}{L} \\ k_{max} &= \frac{2\pi}{2\Delta x} = \frac{\pi}{\Delta x} \end{aligned} \tag{8}$$

Inserting the wave number

$$k_m = m k_{min} = m\frac{\pi}{L} = m\frac{\pi}{N\Delta x} \tag{9}$$

(N is number of cells) into Eq. 6 gives

$$\varepsilon_i^n = \sum_{m=-N}^{N} E_m^n e^{\jmath m i \pi / N} \tag{10}$$

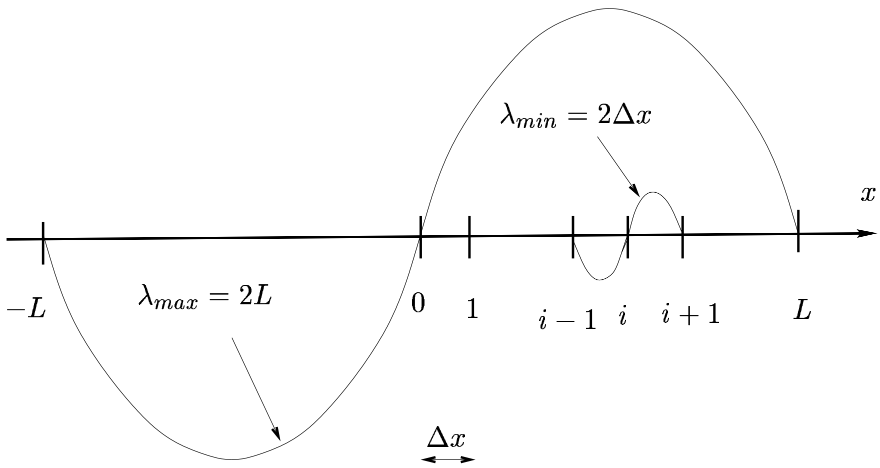

Figure 1: Minimum and maximum wave length and wave number on a computational grid.

Since the problem is linear we can choose *one* Fourier component so that Eq. 5 gives (dropping subscript m)

$$\frac{E^{n+1}e^{\jmath\phi i} - E^n e^{\jmath\phi i}}{\Delta t} + \frac{a}{2\Delta x}\left(E^n e^{\jmath\phi(i+1)} - E^n e^{\jmath\phi(i-1)}\right) = 0 \tag{11}$$

where $\phi = m\pi/N$. Note that low ϕ corresponds to low frequencies and vice versa. The value $\phi = \pi$ corresponds to the highest resolved frequency, i.e. $\lambda = 2\Delta x$, see Fig. 1. Divide by $e^{\jmath\phi i}$ so that

$$E^{n+1} - E^n + \frac{a\Delta t E^n}{2\Delta x}\left(e^{\jmath\phi} - e^{-\jmath\phi}\right) = 0. \tag{12}$$

Using

$$e^{\jmath\phi} = \cos(\phi) + \jmath\sin(\phi) \tag{13}$$

we obtain

$$\begin{aligned} E^{n+1} - E^n + E^n \jmath\sigma\sin(\phi) &= 0 \\ \sigma &= \frac{a\Delta t}{2\Delta x} \end{aligned} \tag{14}$$

Defining the amplification factor

$$G \equiv \frac{E^{n+1}}{E^n} \tag{15}$$

we can write Eq. 14 as

$$G = 1 - \jmath\sigma\sin\phi \tag{16}$$

The condition for a scheme to be stable is that the modulus of G should be smaller than one. We find that that is never true because

$$|G|^2 = 1 + \sigma^2 \sin^2 \phi \geq 1 \tag{17}$$

for all ϕ and σ (σ must be > 0). Thus the scheme is *unstable*. Note that for the highest frequency ($\Phi = \pi$) the damping is neutral ($G = 1$), which means that odd-even oscillations are not affected at all.

3 Second derivative as dissipation term

Assume that we have the equation

$$\frac{\partial T}{\partial t} + \frac{\partial f}{\partial x} = 0 \tag{18}$$

We add a second derivative on the RHS as a dissipation term and analyze the time derivative and the RHS with von Neumann analysis. Thus we have

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}. \tag{19}$$

Discretize the RHS as

$$\frac{\partial^2 T}{\partial x^2} = \frac{T_{i+1} - 2T_i + T_{i-1}}{(\Delta x)^2} \tag{20}$$

and introduce ε as in Section 2

$$\frac{E^{n+1} e^{\jmath\phi i} - E^n e^{\jmath\phi i}}{\Delta t} = \frac{\alpha E^n}{(\Delta x)^2} \left(e^{\jmath\phi(i+1)} - 2e^{\jmath\phi i} + e^{\jmath\phi(i-1)} \right). \tag{21}$$

Divide by $e^{\jmath\phi i}$ and use Eq. 13 so that

$$E^{n+1} - E^n = \frac{\alpha E^n \Delta t}{(\Delta x)^2} \left(e^{\jmath\phi} - 2 + e^{-\jmath\phi} \right) = \frac{\alpha E^n \Delta t}{(\Delta x)^2} \left(2\cos\phi - 2 \right). \tag{22}$$

Finally we get by use of Eq. 15

$$\begin{aligned} G &= 1 - 2\beta(1 - \cos\phi) = 1 - 4\beta \sin^2(\phi/2) \\ \beta &= \frac{\alpha \Delta t}{(\Delta x)^2}. \end{aligned} \tag{23}$$

Provided $\beta \leq 1/4$, we see that $|G| \leq 1$ for all ϕ, i.e. the scheme is dissipative (damping). A central difference of a second derivative can thus be used as a dissipation term. It is not often used as a such, because it would then interfere with the physical diffusion term. However, in compressible flow simulations is is sometimes used as a dissipation term at shocks where it is added locally.

4 Fourth-order derivative as dissipation term

The Rhie-Chow interpolation is equivalent to adding a fourth-order derivative term of the pressure. Below it is shown that this adds dissipation, i.e. it is stabilizing. Assume that we have the equation

$$\frac{\partial T}{\partial t} + \frac{\partial f}{\partial x} = 0 \tag{24}$$

We add a fourth derivative on the LHS (see p. 224, and Eqs. 17.3.16 and 17.3.17 in [1]) as a dissipation term and analyze the time derivative and the RHS with von Neumann analysis. Thus we have

$$\frac{\partial T}{\partial t} = -\eta \frac{\partial^4 T}{\partial x^4}, \tag{25}$$

where η is a small coefficient ($\eta \ll 1$). Discretizing the equation above using explicit discretization in time and central differencing in space gives

$$\frac{T_i^{n+1} - T_i^n}{\Delta t} = -\eta \frac{T_{i-2}^n - 4T_{i-1}^n + 6T_i^n - 4T_{i+1}^n + T_{i+2}^n}{(\Delta x)^4}. \tag{26}$$

Introduce ε as in Section 2 so that

$$\frac{E^{n+1} e^{\jmath\phi i} - E^n e^{\jmath\phi i}}{\Delta t} = -\frac{\eta \Delta t E^n}{(\Delta x)^4}\left(e^{\jmath\phi(i+2)} - 4e^{\jmath\phi(i+1)} + 6e^{\jmath\phi i} - 4e^{\jmath\phi(i-1)} + e^{\jmath\phi(i-2)}\right). \tag{27}$$

Divide by $e^{\jmath\phi i}$ and use Eq. 15

$$G - 1 = -\frac{\eta\Delta t}{(\Delta x)^4}\left(e^{2\jmath\phi} - 4e^{\jmath\phi} + 6 - 4e^{-\jmath\phi} + e^{-2\jmath\phi}\right). \tag{28}$$

Using Eq. 13 gives

$$G = 1 - \frac{\eta\Delta t}{(\Delta x)^4}\left(2\cos(2\phi) - 8\cos\phi + 6\right). \tag{29}$$

If $\eta\Delta t/(\Delta x)^4$ is small (to keep $G > -1$), we find that for high ϕ ($\phi \to \pi$), i.e. high frequencies, the term is indeed dissipative ($|G| < 1$). Please recall that $\phi = \pi$ corresponds to odd-even grid oscillations which are exactly the ones we want to damp.

5 Crank-Nicolson Time Discretization

Discretizing the following equation

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial T^2}{\partial x^2} \tag{30}$$

gives [2]

$$
\begin{aligned}
T_i^{n+1} - T_i^n &= \frac{\sigma}{2}[(T_{i+1}^{n+1} - 2T_i^{n+1} + T_{i+1}^{n+1}) \\
&+ (T_{i+1}^n - 2T_i^n + T_{i+1}^n)] \\
\sigma &= \frac{\alpha \Delta t}{\Delta x^2}
\end{aligned}
\tag{31}
$$

Introduce ε as in Section 2 gives

$$
\begin{aligned}
E^{n+1} - E^n &= \frac{\sigma}{2}[E^{n+1}(e^{-\jmath\phi} - 2 + e^{\jmath\phi}) \\
&+ E^n(e^{-\jmath\phi} - 2 + e^{\jmath\phi})].
\end{aligned}
\tag{32}
$$

Introducing $G = E^{n+1}/E^n$ and using Eq. 13 we get

$$
G - 1 = \frac{\sigma}{2}[G(2\cos\phi - 2) + (2\cos\phi - 2)] \tag{33}
$$

so that

$$
G = \frac{1 + \sigma(\cos\phi - 1)}{1 - \sigma(\cos\phi - 1)} = \frac{1 - 2\sigma\sin^2(\phi/2)}{1 + 2\sigma\sin^2(\phi/2)}. \tag{34}
$$

We see that $|G| < 1$ for all frequencies (all ϕ) which means that the scheme is *stable*.

6 Convergence of the Gauss-Seidel Method (the Scarborough condition)

The Scarborough condition states that [2]

$$
\frac{\sum |a_{nb}|}{a_P} \leq 1 \tag{35}
$$

Below, we will prove this statement.

We want to solve the 2D equation

$$
\begin{aligned}
a_P T_P &= a_E T_E + a_W T_W + a_S T_S + a_N T_N \Rightarrow \\
T_P &= (a_E T_E + a_W T_W + a_S T_S + a_N T_N)/a_P
\end{aligned}
\tag{36}
$$

If we do this by starting in the south-west corner of the computation domain (low i, low j), T_W and T_S will be new values (iteration level $k+1$), whereas T_E and T_N will be old values (iteration level k). We can then rewrite Eq. 36 as

$$
T_P^{k+1} = (a_E T_E^k + a_W T_W^{k+1} + a_S T_S^{k+1} + a_N T_N^k)/a_P. \tag{37}
$$

Now introduce the error

$$\varepsilon^k = T^k - \bar{T}, \tag{38}$$

where $\bar{T}$ is the exact solution to Eq. 36. When solving Eq. 36 iteratively with Gauss-Seidel, we want the error ε to go to zero. Insert Eq. 38 into Eq. 37 so that

$$\varepsilon_P^{k+1} = \frac{a_E}{a_P}\varepsilon_E^k + \frac{a_W}{a_P}\varepsilon_W^{k+1} + \frac{a_S}{a_P}\varepsilon_S^{k+1} + \frac{a_N}{a_P}\varepsilon_N^k. \tag{39}$$

Now let

$$\varepsilon = \max\{|\varepsilon_E^k|, |\varepsilon_W^{k+1}|, |\varepsilon_S^{k+1}|, |\varepsilon_N^k|\}, \tag{40}$$

and insert it into Eq. 39 we get

$$|\varepsilon_P^{k+1}| \leq \varepsilon \frac{|a_E| + |a_W| + |a_S| + |a_N|}{|a_P|}. \tag{41}$$

Thus we find that if the sum of off-diagonal coefficients is smaller than or equal to the diagonal element,

$$\frac{|a_E| + |a_W| + |a_S| + |a_N|}{|a_P|} \leq 1, \tag{42}$$

(the *row condition* or the *Scarborough condition*), then the maximum error will decrease, i.e. the iterative solver will converge. Note that this is a *sufficient* condition but not necessary, i.e. if the Scarborough condition is satisfied, the iterative solver will converge, but the solver may converge even if the condition is not satisfied.

6.1 The Spectral Radius

The condition in Eq. 42 is commonly expressed by saying that the *spectral radius* (the maximum eigenvalue) must be smaller than or equal to one. Eq. 36 can be written on matrix form as

$$AT = B \tag{43}$$

where

$$A = \left[\begin{array}{ccccccccc}
a_{P,2,2} & -a_{E,2,2} & \cdots & -a_{N,2,2} & \cdots & & & & \\
-a_{W,3,2} & a_{P,3,2} & -a_{E,3,2} & \cdots & -a_{N,3,2} & \cdots & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & & \\
\cdots & -a_{S,k,k} & \cdots & -a_{W,k,k} & a_{P,k,k} & -a_{E,k,k} & \cdots & -a_{N,k,k} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & & \\
 & & & & \cdots & -a_{S,ni-1,nj-1} & \cdots & -a_{W,ni-1,nj-1} & a_{P,ni-1,nj-1}
\end{array}\right] \tag{44}$$

Split the matrix A into a lower triangular, a diagonal and an upper triangular matrix as

$$A = L + D + U \tag{45}$$

The Gauss-Seidel solver can now be written on matrix form (cf. Eqs. 37 and 44)

$$(D + L)T^{n+1} = B^n - UT^n. \tag{46}$$

Define the error vector as (cf. Eq. 38)

$$\varepsilon^{n+1} = T^{n+1} - \bar{T} \tag{47}$$

where $\bar{T}$ is the exact solution to the equation system 43. Subtracting the equation

$$A\bar{T} = B \tag{48}$$

from Eq. 46 gives

$$(D + L)\varepsilon^{n+1} = -U\varepsilon^n \Rightarrow \varepsilon^{n+1} = -(D + L)^{-1}U\varepsilon^n. \tag{49}$$

Replace U by Eq. 45, and we get

$$\varepsilon^{n+1} = -(D + L)^{-1}[A - (L + D)]\varepsilon^n = [1 - (D + L)^{-1}A]e^n. \tag{50}$$

Now define $G = 1 - (D + L)^{-1}A$ so that Eq. 50 can be written

$$\varepsilon^{n+1} = Ge^n. \tag{51}$$

If we diagonalized the matrix G by solving the eigenvalue problem

$$|G - I\lambda| = 0 \tag{52}$$

where λ is the eigenvalue vector (with eigenvalues λ_i) of G, Eq. 51 can be written as

$$\varepsilon^{n+1} = \lambda e^n. \tag{53}$$

The condition for convergence (cf. Eq. 42), i.e. the maximum error should decrease $[\max(\varepsilon^{n+1}) \leq \max(\varepsilon^n)]$, can now be expressed as the

$$\max(\lambda_i) \leq 1, \tag{54}$$

i.e. the maximum eigenvalue must be smaller than one. As mentioned above, the maximum eigenvalue is called the *spectral radius.*

References

[1] C. Hirsch. *Numerical Computation of Internal and External Flows: Computational Methods for Inviscid and Viscous Flows*, volume 2. John Wiley & Sons, Chichester, UK, 1990.

[2] H.K. Versteegh and W. Malalasekera. *An Introduction to Computational Fluid Dynamics - The Finite Volume Method.* Longman Scientific & Technical, Harlow, England, 1995.